The University of Chicago Computer Science Department
PRESENTS:

"On the Foundations of Deep Learning: SGD, Overparametrization, and Generalization"

Jason Lee
University of Southern California

Abstract:
We provide new results on the effectiveness of SGD and overparametrization in deep learning.

a) SGD: We show that SGD converges to stationary points for general nonsmooth, nonconvex functions, and that stochastic subgradients can be efficiently computed via Automatic Differentiation. For smooth functions, we show that gradient descent, coordinate descent, ADMM, and many other algorithms, avoid saddle points and converge to local minimizers. For a large family of problems including matrix completion and shallow ReLU networks, this guarantees that gradient descent converges to a global minimum.

b) Overparametrization: We show that gradient descent finds global minimizers of the training loss of overparametrized deep networks in polynomial time.

c) Generalization: For general neural networks, we establish a margin-based theory. The minimizer of the cross-entropy loss with weak regularization is a max-margin predictor, and enjoys stronger generalization guarantees as the amount of overparametrization increases.

d) Algorithmic and Implicit Regularization: We analyze the implicit regularization effects of various optimization algorithms on overparametrized networks. In particular we prove that for least squares with mirror descent, the algorithm converges to the closest solution in terms of the bregman divergence. For linearly separable classification problems, we prove that the steepest descent with respect to a norm solves SVM with respect to the same norm. For over-parametrized non-convex problems such as matrix sensing or neural net with quadratic activation, we prove that gradient descent converges to the minimum nuclear norm solution, which allows for both meaningful optimization and generalization guarantees.

Bio:
Jason Lee is an assistant professor of Data Sciences and Operations in Marshall School of Business, Computer Science Department (courtesy), and a member of the Machine Learning Center. Previously, he was a postdoc in the Computer Science Department at UC Berkeley working with Michael Jordan, and also collaborated with Ben Recht. Prior to that, he was a PhD student in Computational Mathematical Engineering at Stanford University advised by Trevor Hastie and Jonathan Taylor. He received a BS in Mathematics from Duke University advised by Mauro Maggioni. His research interests are broadly in Statistics (Statistical Learning and High-dimensional Statistics), Machine Learning, and Optimization.

Monday, February 25, 2019
10:30 am
JCL 390
Host: Rebecca Willett